VERIFICATION OF TRANSLATION

I, Sabine Frieda Katharina Town of Waldstraße 45,
82386 Oberhausen, Federal Republic of Germany

hereby declare that I am fluent in German and English, am a competent translator from German into English and certify that the attached is a true and accurate translation made by me into the English language of the accompanying officially certified copy of the documents in connection with the Patent Application No. P 44 18 222.8 filed in the Patent Office of the Federal Republic of Germany on the 25th May 1994.

Sabine F.K. Town
15th February 1996

FEDERAL REPUBLIC OF GERMANY

Certificate

The Biopharm Gesellschaft zur biotechnologischen Entwicklung von Pharmaka mbH in 69115 Heidelberg has on the 25th May 1994 filed an application for a patent with the German Patent Office entitled

"New growth/differentiation factor of the TGF-ß family"

and declares that it claims the internal priority for this of the application in the Federal Republic of Germany dated 10th August 1993, file number P 43 26 829.3.

The attached documents are a correct and accurate reproduction of the original documents of this patent application.

The application received the provisional International Patent Classification symbols C 07 K 13/00, C 07 H 21/04, C 07 K 15/28 and A 61 K 37/02 in the German Patent Office.

Munich, 22th August 1994

The President of the German Patent Office

By order of

Lissner

File number: P 44 18 222.8

Biopharm Gesellschaft zur
biotechnologischen Entwicklung
von Pharmaka mbH
Czerny-Ring 22
69115 Heidelberg

New growth/differentiation factor of the TGF-ß family

DESCRIPTION

The present invention concerns a new growth/differentiation factor of the TGF-ß family and DNA sequences coding therefor.

The TGF-ß family of growth factors which includes BMP-, TGF- and inhibin-related proteins (Roberts and Sporn, Handbook of Experimental Pharmacology 95 (1990), 419-472) is particularly relevant for a wide range of medical treatment methods and applications. These factors are suitable in methods which concern wound-healing and tissue regeneration. Furthermore several members from the TGF-ß family induce tissue growth, in particular growth of bones, and therefore play a crucial role in inducing the development of cartilage and bones.

Wozney (Progress in Growth Factor Research 1 (1989), 267-280) and Vale et al (Handbook of Experimental Pharmacology 95 (1990), 211-248) describe various growth factors such as those which are related to the BMP group (bone morphogenetic proteins) and the inhibin group. The members of these groups show significant structural similarities. The precursor of the protein consists of an amino-terminal signal sequence, a propeptide and a carboxy-terminal sequence of about 110 amino acids that are cleaved from the precursor and constitute the mature protein. In addition their members are defined by an amino acid sequence homology. The mature protein contains the most conserved sequences, in particular seven cysteine residues which are conserved among the family members. The TGF-ß-like proteins are multifunctional, hormonally active growth factors. They also have related biological activities such as

chemotactic attraction of cells, promotion of cell differentiation and tissue-inducing capabilities such as cartilage-inducing and bone-inducing capabilities. The US Patent No. 5,013,649 discloses DNA sequences that code for osteo-inductive proteins that are denoted BMP-2 and the US Patent Applications series No. 179 101 and 170 197 disclose the BMP proteins BMP-1 and BMP-3. Moreover many types of cells are able to synthesize TGF-ß-like proteins and practically all cells have TGF-ß receptors.

On the whole these proteins show differences in their structure which leads to significant variations in their exact biological function. In addition they are found in a wide range of different types of tissue and at various stages of development. As a result they can exhibit differences with regard to their exact function e.g. the required cellular physiological environment, their life-span, their target sites, their requirements for auxiliary factors and their stability against degradation. Thus, although a multitude of proteins have been described that exhibit a tissue-inductive and in particular osteo-inductive potential, their natural functions in the organism and - more significantly - their medical relevance still have to be investigated in detail. It is thought to be highly probable that members of the TGF-ß family are present that are still unknown which are important for osteogenesis or the differentiation/induction of other types of tissue. A major difficulty in the isolation of these new TGF-ß-like proteins is, however, that their functions cannot yet be described exactly enough to develop highly discriminating bioassays. On the other hand the expected nucleotide sequence homology to other members of the family is too low to enable a screening by classical

nucleic acid hybridization techniques. Nevertheless the further isolation and characterization of new TGF-ß-like proteins is urgently required in order to provide further inducing and differentiation proteins that fulfil all the desired medical requirements. These factors could be used medically in the healing of lesions and the treatment of degenerative diseases of bones and/or other types of tissue such as the kidney or the liver.

A nucleotide and amino acid sequence for the TGF-ß protein MP-52 is stated in the Patent Application PCT/EP93/00350 in which the sequence corresponding to the mature peptide and a major portion of the sequence corresponding to the propeptide of MP-52 are given. The complete sequence of the propeptide MP-52 is not disclosed.

The object on which the present invention is based is to provide DNA sequences that code for new members of the TGF-ß protein family with mitogenic and/or differentiation-inductive e.g. osteo-inductive potential. The object of the present invention was therefore in particular to provide the complete DNA and amino acid sequence of the TGF protein MP-52.

This object is achieved by a DNA molecule that codes for a protein of the TGF-ß family and which comprises
(a) the part coding for the mature protein and if desired further functional parts of the nucleotide sequence shown in SEQ ID NO. 1,
(b) a nucleotide sequence corresponding to a sequence from (a) within the scope of the degeneracy of the genetic code,

(c) an allelic derivative of a nucleotide sequence corresponding to one of the sequences from (a) and (b) or
(d) a sequence hybridizing with one of the sequences from (a), (b) or (c)
provided that a DNA molecule according to (d) contains at least the part coding for a mature protein from the TGF-ß family.

Further embodiments of the present invention concern the subject matter of claims 2 to 10. Other features and advantages of the invention can be derived from the description of the preferred embodiments and figures. The sequence protocols and figures are now briefly described.

SEQ ID NO. 1 shows the complete nucleotide sequence of the DNA coding for the TGF-ß protein MP-52. The ATG start codon starts with nucleotide 640. The start of the mature protein begins after nucleotide 1782.

SEQ ID NO. 2 shows the complete amino acid sequence of the TGF-ß protein MP-52 which was derived from the nucleotide sequence shown in SEQ ID NO. 1.

Figure 1 shows a comparison between the amino acid sequence of MP-52 and several members of the BMP protein family starting with the first of the seven conserved cysteine residues. * denotes that the amino acid is the same in all compared proteins; + denotes that the amino acid corresponds in at least one of the proteins compared to MP-52.

Figure 2 shows the nucleotide sequences of the oligonucleotide primers that were used in the present invention and a comparison of these sequences with known members of the TGF-ß family. M denotes A or C, S denotes C or G, R denotes A or G and K denotes G or T. 2a shows the sequence of the primer OD, 2b shows the sequence of the primer OID.

The present invention encompasses at least the part coding for the mature protein and if desired further functional parts of the nucleotide sequence shown in SEQ ID NO.1 as well as sequences that correspond to this sequence within the scope of the degeneracy of the genetic code and allelic derivatives of such sequences. In addition the present invention also encompasses sequences that hybridize with such sequences provided that such a DNA molecule completely contains at least the part coding for the mature protein of the TGF-ß family.

The term "functional part" within the sense of the present invention denotes a protein part which is capable of acting for example as a signal peptide, propeptide or as a mature protein part i.e. it fulfils at least one of the biological functions of the natural protein parts of MP-52.

The region coding for the mature part of the protein extends from nucleotides 1783 - 2142 of the sequence shown in SEQ ID NO.1. If desired, the DNA molecule can also comprise further functional parts of the sequence shown in SEQ ID NO. 1, namely the nucleotide sequences coding for the signal or/and propeptide part. It is particularly preferred that the DNA molecule comprises

the sequence for the signal and the propeptide part and the part of the mature protein i.e. nucleotides 640-2142 of the sequence shown in SEQ ID NO. 1. On the other hand the DNA molecule can also comprise functional signal or/and propeptide parts from other proteins in addition to the part coding for the mature protein, in particular from other proteins of the TGF-ß family e.g. the above-mentioned BMP proteins. The respective nucleotide sequences are given in the references mentioned above to the disclosure of which reference is hereby being made.

Moreover the present invention also encompasses a DNA molecule as defined above that contains a non-coding intron sequence between nucleotides 1270 and 1271 of the sequence shown in SEQ ID NO. 1. This intron sequence is contained in the plasmid SKL 52 (H3) MP12 which is deposited at the DSM and has the genomic nucleic acid sequence of MP-52.

The invention also encompasses the cDNA sequence of the MP-52 protein coded by the phage λ 15.1. This sequence starts with nucleotide 321 of SEQ ID NO. 1.

Although the allelic, degenerate and hybridizing sequences which are encompassed by the present invention have structural differences due to slight changes in their nucleotide or/and amino acid sequence, the proteins coded by such sequences still essentially have the same useful properties that enable their use in basically the same medical applications.

The term "hybridization" according to the present invention means the usual hybridization conditions, preferably conditions with a salt concentration of 6 x

SSC at 62 to 66°C followed by a one hour wash with 0.6 x SSC, 0.1 % SDS at 62 to 66°C. It is particularly preferred that the term "hybridization" denotes stringent hybridization conditions with a salt concentration of 4 x SSC at 62 to 66°C followed by a one hour wash with 0.1 x SSC, 0.1 % SDS at 62 to 66°C.

Preferred embodiments of the present invention are DNA sequences as defined above that are obtainable from vertebrates, preferably mammals such as pigs, cows and rodents such as rats or mice and in particular from primates such as humans.

A particularly preferred embodiment of the present invention is the sequence denoted MP-52 shown in SEQ ID NO. 1. The transcripts of MP-52 were obtained from embryonic tissue and code for a protein which has a considerable amino acid homology to the mature portion of BMP-like proteins (see Fig. 1). The protein sequences of BMP2 (=BMP2A) and BMP4 (=BMP2B) are described by Wozney et al., Science 242 (1988), 1528-1534. The corresponding sequences of BMP5, BMP6 and BMP7 are described by Celeste et al., Proc. Natl. Acad. Sci. USA 87 (1990), 9843-9847. Several typical sequence homologies which are specific for known BMP sequences have also been found in the propeptide part of MP-52 whereas other parts of the precursor part of MP-52 exhibit considerable differences to BMP precursors.

In addition the present invention concerns a vector that contains at least one copy of a DNA molecule according to the invention. The DNA sequence according to the invention is preferably operatively linked to an expression control sequence in such a vector. Such

vectors are suitable for the production of TGF-ß-like proteins in stably or transiently transformed cells. Various animal, plant, fungal and bacterial systems can be used for the transformation and subsequent culture. The vectors according to the invention preferably contain sequences necessary for replication in the host cell and are autonomously replicable. Furthermore the use of vectors is preferred that contain selectable marker genes which can be used to detect transformation of a host cell.

In addition the invention concerns a host cell which is transformed with a DNA according to the invention or with a vector according to the invention. Examples of suitable host cells include various eukaryotic and prokaryotic cells such as E. coli, insect cells, plant cells, mammalian cells and fungi such as yeast.

In addition the invention concerns a protein of the TGF-ß family which is coded by a DNA sequence according to claim 1. The protein according to the invention preferably has the amino acid sequence shown in SEQ ID NO. 2 or if desired functional parts thereof and exhibits biological properties such as tissue-inductive in particular osteo-inductive or/and mitogenic capabilities that may be relevant for a therapeutic application. The above-mentioned characteristics of the protein can vary depending on the formation of homodimers or heterodimers. Such structures may also prove to be suitable for clinical applications.

The biological properties of the proteins according to the invention, in particular the mitogenic and osteo-inductive potential can be determined for example in

assays according to Roberts et al., PNAS 78 (1981), 5339-5343, Seyedin et al., PNAS 82 (1985), 2267-2271 or Sampath and Reddi, PNAS 78 (1981), 7599-7603.

Furthermore the present invention concerns a process for the production of a protein of the TGF-ß family which is characterized in that a host cell transformed with a DNA according to the invention or with a vector according to the invention is cultured and the TGF-ß protein is isolated from the cell or/and culture supernatant. Such a process comprises culturing the transformed host cell in a suitable culture medium and purifying the TGF-ß-like protein produced. In this manner the process enables the production of an adequate amount of the desired protein for use in medical treatment or in applications using cell culture techniques which require growth factors. The host cell can be a bacterium such as Bacillus or E. coli, a fungus such as yeast, a plant cell such as tobacco, potato or arabidopsis or an animal cell and in particular a vertebrate cell line such as MoCOS or CHO cell lines or an insect cell line.

Yet another subject matter of the present invention is the provision of pharmaceutical compositions which contain a pharmaceutically active amount of a TGF-ß-like protein according to the invention as the active substance. If desired, such a composition comprises a pharmaceutically acceptable carrier substance, auxiliary substance, diluent or filler. Such a pharmaceutical composition can be used in wound-healing and tissue regeneration as well as in the healing of damage to bones, cartilage, connective tissue, skin, mucous membranes, epithelium or teeth and in dental implants either alone or in combination with other active substances e.g. other proteins of the TGF-ß family or

growth factors such as EGF (epidermal growth factor) or PDGF (platelet derived growth factor). Moreover such a pharmaceutical composition can be used in the prevention of diseases such as the prevention of osteoporosis and arthrosis.

Another possible clinical application of the TGF-ß protein according to the invention is to use it as a suppressor of an immunoreaction to prevent rejection of organ transplants or its application in connection with angiogenesis. The pharmaceutical composition according to the invention can also be used prophylactically or in cosmetic surgery. In addition administration of the composition is not limited to humans but can also encompass animals and in particular pets.

Finally the present invention also concerns an antibody that binds specifically to the proteins according to the invention or such an antibody fragment (e.g. Fab or Fab'). Processes for the production of such a specific antibody or antibody fragment are part of the general technical knowledge of an average person skilled in the art. Such an antibody is preferably a monoclonal antibody. Such antibodies or antibody fragments may also be suitable for diagnostic methods.

It is intended to elucidate the invention further by the following example.

Example 1

Isolation of MP-52

1.1 Total RNA was isolated from human embryonic tissue (8 to 9 weeks old) according to the method of Chirgwin et al., Biochemistry 18 (1979), 5294-5299. Poly(A+) RNA was separated from the total RNA by oligo (dT) chromatography according to the manufacturer's instructions (Stratagene Poly (A) Quick columns).

1.2 For the reverse transcription reaction 1 to 2.5 µg poly (A+) RNA was heated for 5 minutes to 65°C and quickly cooled on ice. The reaction mixture contained 27 U RNA-Guard (Pharmacia), 2.5 µg oligo (dT)12-18 (Pharmacia), 5 x buffer (250 mmol/l Tris/HCl, pH 8.5, 50 mmol/l $MgCl_2$, 50 mmol/l DTT, 5 mmol/l of each dNTP, 600 mmol/l KCl) and 20 U AMV reverse transcriptase (Boehringer Mannheim) per µg poly (A+) RNA. The reaction mixture (25 µl) was incubated for 2 hours at 42°C.

1.3 The deoxynucleotide primers OD and OID shown in Fig. 2 were prepared on an automatic DNA synthesizer (Biosearch). The purification was carried out by denaturing polyacrylamide gel electrophoresis and isolating the main bands from the gel by means of isotachophoresis. The oligonucleotides were designed by comparing the nucleic acid sequences of known members of the TGF-ß family and selecting regions with the highest conservation. A comparison of this region is shown in Fig. 2. In order to facilitate cloning both nucleotides contained EcoRI restriction sites and

OD additionally contained a NcoI restriction site at its 5' terminus.

1.4 cDNA corresponding to 20 ng poly (A+) RNA was used as the starting material in the PCR reaction. The reaction was carried out in a volume of 50 µl and contained 1 x PCR buffer (16.6 mmol/l $(NH_4)_2SO_4$, 67 mmol/l Tris/HCl pH 8.8, 2 mmol/l $MgCl_2$, 6.7 µmol/l EDTA, 10 mmol/l ß-mercapto-ethanol, 170 µg/ml bovine serum albumin (Gibco), 200 µmol/l of each dNTP (Pharmacia), 30 pmol of each oligonucleotide (OD and OID) and 1.5 U Taq polymerase (AmpliTaq, Perkin Elmer Cetus). The reaction mixture was overlayed with paraffin and 40 PCR cycles were carried out. The products of the PCR reaction were purified by phenol/chloroform extraction and concentrated by ethanol precipitation.

1.5 The PCR reaction product was cleaved with the restriction enzymes SphI (Pharmacia) and AlwNI (Biolabs) according to the manufacturer's instructions.

1.6 The products of the restriction cleavage were fractionated by Agarose gel electrophoresis. After staining with ethidium bromide, uncleaved amplification products were cut out of the gel and isolated by phenol extraction. The DNA obtained was subsequently purified twice by phenol/chloroform extraction.

1.7 After an ethanol precipitation, a quarter or a fifth of the isolated DNA was reamplified using the same conditions as for the primary amplification except that the number of cycles was reduced to 13. The reamplification products were purified, cleaved with the same enzymes as above and the uncleaved products were isolated from Agarose gels as elucidated above for the amplification products. The reamplification step was repeated twice.

1.8 After the last isolation from the gel, the amplification products were cleaved by 4 units EcoRI (Pharmacia) under the conditions recommended by the manufacturer. A fourth of the restriction mixture was ligated into the vector pBluescriptII SK+ (Stratagene) cleaved with EcoRI. 24 clones were analysed further by means of sequencing after ligation. The sample cleaved with AlwNI and SphI resulted in a new sequence that was denoted MP-52. The other clones mainly contained BMP6 sequences and one contained a BMP7 sequence.

The clone was completed up to the 3' end of the cDNA according to the method described in detail by Frohmann (Amplifications, published by Perkin-Elmer Corp., Issue 5 (1990), pp 11-15). The same embryonic mRNA that had been used to isolate the first fragment of MP-52 was reversally transcribed as described above. The amplification was carried out using the adapter primer (AGAATTCGCATGCCATGGTCGACG) and an inner primer (CTTGAGTACGAGGCTTTCCACTG) of the MP-52 sequence. The amplification products were reamplified using an overlapping adapter primer (ATTCGCATGCCATGGTCGACGAAG) and an overlapping internal primer (GGAGCCCACGAATCATGCAGTCA) of the MP-52 sequence. After

restriction cleavage with NcoI the reamplification products were cloned and sequenced into a vector that was cleaved in the same way (pUC 19 (Pharmacia No. 27-4951--01) having a modified multiple cloning site which contains a single NcoI restriction site) and sequenced. The clones were characterized by their sequence overlapping at the 3' end of the known MP-52 sequence. One of these was used as a probe to screen a human genomic gene bank (Stratagene No. 946203) according to a method described in detail by Ausubel et al. (Current Protocols in Molecular Biology, published by Greene Publishing Associates and Wiley-Interscience (1989)). One phage (λ2.7.4) was isolected from 8 x 10^5 λ phages which contained an insertion of about 20 kb and which is deposited at the DSM under the depository number 7387. This clone contains further sequence informations at the 5' end in addition to the sequence isolated from mRNA by the described amplification methods.

For the sequence analysis a HindIII fragment of about 7.5 kb was subcloned into a vector cleaved in the same manner (Bluescript SK, Stratagene No. 212206). This plasmid denoted SKL 52 (H3) MP12 was also deposited at the DSM under the depository number 7353. The sequence information shown in SEQ ID NO. 1 was derived from the phage λ 2.7.4.. The ATG at position 640 is the first ATG within the reading frame (a stop codon occurs at position 403). Based on the sequence data it may be assumed that this is the start codon for the translation.

The genomic DNA contains an intron of about 2 kb between base pairs 1270 and 1271 of SEQ ID NO: 1. The sequence of the intron is not shown. The correctness of the splicing site was confirmed by sequencing an

amplification product which was derived from cDNA containing this region. These sequence informations were obtained using a slightly modified method which is described in detail by Frohmann (Amplifications, published by Perkin-Elmer Corporation, Issue 5 (1990), pp 11-15). The same embryonic RNA that was also used to isolate the 3' end of MP-52 was reversally transcribed using an internal primer orientated in the 5' direction of the MP-52 sequence (ACAGCAGGTGGGTGGTGTGGACT). A polyA tail was attached to the 5' end of the first cDNA strand using terminal transferase. A two-step amplification was carried out, firstly by using a primer composed of oligo dT and an adapter sequence (AGAATTCGCATGCCATGGTCGACGAAGC(T16)) and secondly an adapter primer (AGAATTCGCATGCCATGGTCGACG) and an internal primer (CCAGCAGCCCATCCTTCTCC) from the MP-52 sequence. The amplification products were reamplified using the same adapter primer and an overlapping internal primer (TCCAGGGCACTAATGTCAAACACG) from the MP-52 sequence. Subsequently the reamplification products were reamplified using an overlapping adapter primer (ATTCGCATGCCATGGTCGACGAAG) and an overlapping internal primer (ACTAATGTCAAACACGTACCTCTG) from the MP-52 sequence. The final reamplification products were cloned with blunt ends into a vector (Bluescript SK, Stratagene No. 212206) which had been cleaved with EcoRV. The clones were characterized by their sequence overlapping with the DNA of λ 2.7.4..

In addition a cDNA bank - produced from RNA of fibroblasts and cloned into λgt10 - was screened. In this process 2 x 10^6 phages were tested using a ca. 1 kb fragment of genomic MP-52 DNA (2nd exon up to the HindIII restriction site in the 3' untranslated region) as a radioactive probe. 17 mixed plaques were picked out

which were checked using primers from the 5' and 3' region of the MP-52 sequence. Subsequently 8 phage plaques were selected and isolated. cDNA was isolated from the phage by an EcoRI partial cleavage and cloned into the Bluescript vector that was also cleaved with EcoRI.

Sequencing of one of the resulting plasmids SK52L15.1MP25 showed that the longest phage (15.1) starts at nucleotide No. 321 of SEQ ID NO. 1. In addition the splicing position (nucleotide 1270) was confirmed by the sequencing.

The plasmid SKL 52 (H3) MP12 was deposited on 10th December 1992 at the DSM ("Deutsche Sammlung von Mikroorganismen und Zellkulturen, Mascheroder Weg 1b, 3300 Braunschweig") under the depository number 7353.

The phage λ 2.7.4 was deposited on 13th January 1993 at the DSM under the depository number 7387.

The plasmid SK52L15.1MP25 was deposited on 16th July 1993 at the DSM under the depository number 8421.

Example 2

Expression of MP52

Various systems were checked for the expression of MP52. The use of Vaccinia viruses as an expression system is described in detail and capable of being reproduced by a person skilled in the art in Current Protocols in Molecular Biology (Ausubel et al., Greene Publishing Associates and Wiley-Interscience, Wiley & Sons),

abbreviated CP in the following, in chapter 16 unit 16.15-16.18. The system is based on the fact that foreign DNA can be integrated by homologous recombination into the genome of the Vaccinia virus using certain vectors. For this purpose the vector used contains the TK (thymidine kinase) gene from the Vaccinia genome. In order to enable selection for recombinant viruses the vector in addition contains the E. coli xanthine-guanine phosphoribosyl transferase gene (gpt) (Falkner et al., J. Virol. 62 (1988), 1849-1854). The cDNA with the entire region coding for MP52 was cloned into this vector. The cDNA comes from plasmid SK52L15.1MP25 (DSM, depository number 8421) which was, however, firstly deleted and subcloned in order to remove a large portion of the 5' untranslated region. For this the plasmid SK52L15.1MP25 was linearized with SalI and the 5' end was deleted stepwise with the EcoIII/mung bean kit (Stratagene #200330) according to the manufacturer's instructions. After restriction with BamHI, the MP52 cDNAs that had been deleted to different extents were isolated from the residual vector by an Agarose gel and subcloned (pSK52s) according to standard methods (Sambrook et al., Molecular Cloning, second edition, Cold Spring Harbor Laboratory Press 1989) in a pBluescriptII SK vector (Stratagene #212206) restricted with EcoRV and BamHI. All restrictions were carried out according to the manufacturer's instructions. Sequencing with Sequenase (USB/Amersham #70770) yielded inter alia a clone which starts with nucleotide 576 in SEQ ID NO. 1 (64 base pairs distant from the start codon). The cDNA insert was isolated from this by restriction with SalI and SacI and cloned into the likewise cleaved vector for recombination in Vaccinia. The resulting plasmid (pBP1MP52s) was deposited on 24th May 1994 at the DSM (deposit number 9217) and used for the production of

recombinant Vaccinia viruses. For this up to 80 % confluent 143B cells (HuTk, ATCC CRL 8303) in 35 mm culture dishes were infected with Vaccinia wild-type virus in 2 ml PBS for 30 minutes at room temperature while shaking occasionally (1 virus per 10 cells). After aspirating the supernatant and adding 2 ml culture medium (MEM, Gibco BRL #041-01095), it was incubated for 2 hours at 37°C. The medium was subsequently removed and transformation of these cells was achieved with 100 ng pBP1MP52s, 2 µg carrier DNA (calf thymus, Boehringer Mannheim #104175) and 10 µl Lipofectin (Gibco BRL # 18292-011) in 1 ml MEM for 15 hours at 37°C. After addition of 1 ml MEM containing 20 % FCS (Gibco BRL #011-06290), they were incubated for a further 24 hours at 37°C and subsequently the lysed cells were frozen.

The gpt selection for xanthine-guanine phosphoribosyl transferase and isolation and amplification of individual recombinant viruses was essentially carried out as described in unit 16.17 of CP except that RK13 cells (ATCC CCL 37) were used.

Integration of MP52 cDNA into the virus genome was confirmed by dot blot and Southern blot analysis (CP unit 16.18). A recombinant virus was used for expression analyses in the cell line 143B (HuTk-, ATCC CRL 8303, human). Confluent cells were infected for 45 minutes at 37°C with a number of viruses corresponding to the number of cells and subsequently added to the respective culture medium (MEM, Gibco BRL #041-01095) containing 10 % FCS and penicillin/streptomycin (1:500, Gibco BRL # 043-05140H). After 6 hours at 37°C, the medium was removed, the cells were washed twice with e.g. HBSS (Gibco BRL #042-04180M) and production medium (e.g. MEM) without FCS was added. After 20 to 22 hours of

production the cell supernatant was collected. The expression was analysed by means of Western blots according to standard methods (CP unit 10.8). For this the proteins from 100 to 500 µl cell culture supernatant were precipitated by addition of an equivalent volume of acetone and incubating for at least one hour on ice and centrifuged. After resuspending the pellet in application buffer (7 M urea, 1 % SDS, 7 mM sodium dihydrogen phosphate, 0.01 % bromophenol blue and if desired 1 % ß-mercaptoethanol) separation was carried out in 15 % polyacrylamide gels. A prestained protein molecular weight standard (Gibco BRL #26041-020) was used as the marker proteins. Transfer onto a PVDF membrane (Immobilon #IPVH00010) and blocking the membrane were carried out according to standard methods.

In order to detect MP52 on the membrane, polyclonal antibodies against MP52 had been produced in chickens as well as in rabbits. For this the mature part of MP52 with 6 histidines at the N-terminus was expressed in E. coli and purified as described for example in Hochuli et al. (BIO/Technology, Vol. 6, 1321-1325 (1988)). Both antibodies enable the specific detection of expression of MP52 wherein dimeric MP52 is less efficiently recognized than monomeric. Chicken antibodies were used for the Western blot in Figure 3 that had been specifically purified by means of PEG precipitation (Thalley et al., BIO/Technology vol. 8 934-938 (1990)) and by means of membrane-bound antigen (mature MP52 containing 6 histidines) (18.17 in Sambrook et al., Molecular cloning, second edition, Cold Spring Harbor Laboratory Press 1989). Anti-chicken IgG with coupled alkaline phosphatase (Sigma A9171) was used as the second antibody. The detection was carried out using the Tropix Western-Light protein detection kit (Serva

#WL10RC) according to the manufacturer's instructions.

The Western blot in Figure 3 shows that MP52-specific bands only occur in the recombinant viruses but not in the wild-type viruses (without integrated foreign-DNA). The expression of MP52 results in a secreted protein having an apparent molecular weight of about 25 kDa in the gel under non-reducing conditions. The protein migrates in the gel at 14 to 15 kDa under reducing conditions. These results show that MP52 is expressed as a dimeric mature protein. The weak bands appearing in the region above 60 kDa that occur in the Western blot are probably residues of the uncleaved precursor proteins. The migration properties also confirms the theoretical molecular weights that can be derived from SEQ ID NO. 2 according to which mature, monomeric MP52 has a size of 13.6 kDa.

It has been proven to be possible to express MP52 and cleave the precursor protein to mature MP52 in various cell lines. C127 (ATCC CRL 1616, mouse), BHK21 (ATCC CCL 10, hamster), MRC-5 (ATCC CCL 171, human) and 3T6-Swiss albino (ATCC CCL 96, mouse) cells were tested.

Expression and cleavage to form mature MP52 was also demonstrated in a further eukaryotic expression system. For this cDNA from MP52 (starting with nucleotide 576) was cloned into the expression plasmid pSG5 (Stratagene #216201). The plasmid pSK52s was restricted with ClaI and XbaI and the protruding ends of the MP52 insert were made blunt by T4 polymerase treatment. Cloning into the vector pSG5, that had been restricted with EcoRI and likewise blunt ended by T4 polymerase treatment, was carried out according to standard methods. All enzymatic

reactions were carried out according to the instructions of the manufacturer. Correct orientation of the MP52 insert was ensured by restriction analysis and sequencing with the T7 primer (Stratagene #300302). The resulting plasmid pSG52s (deposited on 17.05.94 at the DSM with the deposit number DSM 9204) can be cotransformed with a vector that codes for a selectable marker such as e.g. the gene for G418 resistance in order to obtain stable cell lines. For this purpose pSG52s was cotransformed with the plasmid p3616 (deposited on 17.05.94 at the DSM with the deposit number DSM 9203) in L929 cells (ATCC CCL1, mouse) using Lipofectin (Gibco BRL #18292-011) according to the manufacturer's instructions. Selection with G418 was carried out according to methods familiar to a person skilled in the art (CP, unit 9.5) and it resulted in a cell line that produced detectable mature MP52 in the Western blot.

CLAIMS

1. DNA molecule that codes for a protein of the TGF-ß family and which comprises
(a) the part coding for the mature protein and if desired further functional parts of the nucleotide sequence shown in SEQ ID NO. 1,
(b) a nucleotide sequence corresponding to the sequence from (a) within the scope of the degeneracy of the genetic code,
(c) a nucleotide sequence corresponding to an allelic derivative of one of the sequences from (a) and (b), or
(d) a nucleotide sequence hybridizing with one of the sequences from (a), (b) or (c)
provided that a DNA molecule according to (d) completely contains at least the part coding for a mature protein of the TGF-ß family.

2. Vector,
w h e r e i n
it contains at least one copy of a DNA molecule as claimed in claim 1.

3. Host cell,
w h e r e i n
it is transformed by a DNA as claimed in claim 1 or by a vector as claimed in claim 2.

4. Host cell as claimed in claim 3,
w h e r e i n
it is a bacterium, a fungus, a plant or an animal cell.

5. Protein of the TGF-ß family which is coded by a DNA sequence as claimed in claim 1.

5. Protein as claimed in claim 5,
w h e r e i n
it has the amino acid sequence shown in SEQ ID NO. 2 or, if desired, functional parts thereof.

7. Process for the production of a protein of the TGF-ß family,
w h e r e i n
a host cell as claimed in claim 3 or 4 is cultured and the TGF-ß protein is isolated from the cell or/and from the culture supernatant.

8. Pharmaceutical composition,
w h e r e i n
it contains at least one protein as claimed in claim 5 or 6 as the active substance if desired, together with the usual pharmaceutical carrier substances, auxiliary substances, diluents or fillers.

9. Pharmaceutical composition as claimed in claim 8 for the treatment or prevention of damage to bone, cartilage, connective tissues, skin, mucous membranes, epithelium or teeth, for application in dental implants and for application in wound-healing and tissue regeneration processes.

10. Antibody or antibody fragments,
w h e r e i n
they bind to a protein as claimed in claim 5 or 6.

ABSTRACT

The invention concerns a protein of the TGF-ß family, the DNA coding therefor and a pharmaceutical composition containing the protein.

SEQ ID NO.1
TYPE OF SEQUENCE: nucleic acid sequence
NAME AND ORIGIN: MP-52 gene
LENGTH: 2703 nucleotides

CCATGGCCTCGAAAGGGCAGCGGTGATTTTTTTCACATAAATATATCGCACTTAAATGAG
TTTAGACAGCATGACATCAGAGAGTAATTAAATTGGTTTGGGTTGGAATTCCGTTTCCAA
TTCCTGAGTTCAGGTTTGTAAAAGATTTTTCTGAGCACCTGCAGGCCTGTGAGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTGTGTGAAGTATTTTCACTGGAAAGGATTCAAAACTA
GGGGGAAAAAAAAAACTGGAGCACACAGGCAGCATTACGCCATTCTTCCTTCTTGGAAAAA
TCCCTCAGCCTTATACAAGCCTCCTTCAAGCCCTCAGTCAGTTGTGCAGGAGAAAGGGGG
CGGTTGGCTTTCTCCTTTCAAGAACGAGTTATTTCAGCTGCTGACTGGAGACGGTGCAC
GTCTGGATACGAGAGCATTTCCACTATGGGACTGGATACAAACACACACCCGGCAGACTT
CAAGAGTCTCAGACTGAGGAGAAAGCCTTTCCTTCTGCTGCTACTGCTGCTGCCGCTGCT
TTTGAAAGTCCACTCCTTTCATGGTTTTTCCTGCCAAACCAGAGGCACCTTTGCTGCTGC
CGCTGTTCTCTTTGGTGTCATTCAGCGGCTGGCCAGAGGATGAGACTCCCCAAACTCCTC
ACTTTCTTGCTTTGGTACCTGGCTTGGCTGGACCTGGAATTCATCTGCACTGTGTTGGGT
GCCCCTGACTTGGGCCAGAGACCCCAGGGGACCAGGCCAGGATTGGCCAAAGCAGAGGCC
AAGGAGAGGCCCCCCCCTGGCCCGGAACGTCTTCAGGCCAGGGGGTCACAGCTATGGTGGG
GGGGCCACCAATGCCAATGCCAGGGCAAAGGGAGGCACCGGGCAGACAGGAGGCCTGACA
CAGCCCAAGAAGGATGAACCCAAAAAGCTGCCCCCCAGACCGGGCGGCCCTGAACCCAAG
CCAGGACACCCTCCCCAAACAAGGCAGGCTACAGCCCGGACTGTGACCCCAAAAGGACAG
CTTCCCGGAGGCAAGGCACCCCCAAAAGCAGGATCTGTCCCCAGCTCCTTCCTGCTGAAG
AAGGCCAGGGAGCCCGGGCCCCCACGAGAGCCCAAGGAGCCGTTTCGCCCACCCCCCATC
ACACCCCACGAGTACATGCTCTCGCTGTACAGGACGCTGTCCGATGCTGACAGAAAGGGA
GGCAACAGCAGCGTGAAGTTGGAGGCTGGCCTGGCCAACACCATCACCAGCTTTATTGAC
AAAGGGCAAGATGACCGAGGTCCCGTGGTCAGGAAGCAGAGGTACGTGTTTGACATTAGT
GCCCTGGAGAAGGATGGGCTGCTGGGGGCCGAGCTGCGGATCTTGCGGAAGAAGCCCTCG
GACACGGCCAAGCCAGCGGCCCCCGGAGGCGGGCGGGCTGCCCAGCTGAAGCTGTCCAGC
TGCCCCAGCGGCCGGCAGCCGGCCTCCTTGCTGGATGTGCGCTCCGTGCCAGGCCTGGAC
GGATCTGGCTGGGAGGTGTTCGACATCTGGAAGCTCTTCCGAAACTTTAAGAACTCGGCC
CAGCTGTGCCTGGAGCTGGAGGCCTGGGAACGGGGCAGGGCCGTGGACCTCCGTGGCCTG
GGCTTCGACCGCGCCGCCCGGCAGGTCCACGAGAAGGCCCTGTTCCTGGTGTTTGGCGC
ACCAAGAAACGGGACCTGTTCTTTAATGAGATTAAGGCCCGCTCTGGCCAGGACGATAAG

ACCGTGTATGAGTACCTGTTCAGCCAGCGGCGAAAACGGCGGCGCCCCACTGGCCACTCGC
CAGGGCAAGCGACCCAGCAAGAACCTTAAGGCTCGCTGCAGTCGGAAGGCACTGCATGTC
AACTTCAAGGACATGGGCTGGGACGACTGGATCATCGCACCCCTTGAGTACGAGGCTTTC
CACTGCGAGGGGCTGTGCGAGTTCCCATTGCGCTCCCACCTGGAGCCCACGAATCATGCA
GTCATCCAGACCCTGATGAACTCCATGGACCCCGAGTCCACACCACCCACCTGCTGTGTG
CCCACGCGGCTGAGTCCCATCAGCATCCTCTTCATTGACTCTGCCAACAACGTGGTGTAT
AAGCAGTATGAGGACATGGTCGTGGAGTCGTGTGGCTGCAGGTAGCAGCACTGGCCCTCT
GTCTTCCTGGGTGGCACATCCCAAGAGCCCCTTCCTGCACTCCTGGAATCACAGAGGGGT
CAGGAAGCTGTGGCAGGAGCATCTACACAGCTTGGGTGAAAGGGGATTCCAATAAGCTTG
CTCGCTCTCTGAGTGTGACTTGGGCTAAAGGCCCCCTTTTATCCACAAGTTCCCCTGGCT
GAGGATTGCTGCCCGTCTGCTGATGTGACCAGTGGCAGGCACAGGTCCAGGGAGACAGAC
TCTGAATGGGACTGAGTCCCAGGAAACAGTGCTTTCCGATGAGACTCAGCCCACCATTTC
TCCTCACCTGGGCCTTCTCAGCCTCTGGACTCTCCTAAGCACCTCTCAGGAGAGCCACAG
GTGCCACTGCCTCCTCAAATCACATTTGTGCCTGGTGACTTCCTGTCCCTGGGACAGTTG
AGAAGCTGACTGGGCAAGAGTGGGAGAGAAGAGGAGAGGGCTTGGATAGAGTTGAGGAGT
GTGAGGCTGTTAGACTGTTAGATTTAAATGTATATTGATGAGATAAAAAGCAAAACTGTG
CCT

SEQ ID NO. 2
TYPE OF SEQUENCE: amino acid sequence
NAME AND ORIGIN: MP-52 protein
LENGTH: 501 amino acids

MRLPKLLTFL LWYLAWLDLE FICTVLGAPD LGQRPQGTRP GLAKAEAKER
PPLARNVFRP GGHSYGGGAT NANARAKGGT GQTGGLTQPK KDEPKKLPPR
PGGPEPKPGH PPQTRQATAR TVTPKGQLPG GKAPPKAGSV PSSFLLKKAR
EPGPPREPKE PFRPPPITPH EYMLSLYRTL SDADRKGGNS SVKLEAGLAN
TITSFIDKGQ DDRGPVVRKQ RYVFDISALE KDGLLGAELR ILRKKPSDTA
KPAAPGGGRA AQLKLSSCPS GRQPASLLDV RSVPGLDGSG WEVFDIWKLF
RNFKNSAQLC LELEAWERGR AVDLRGLGFD RAARQVHEKA LFLVFGRTKK
RDLFFNEIKA RSGQDDKTVY EYLFSQRRKR RAPLATRQGK RPSKNLKARC
SRKALHVNFK DMGWDDWIIA PLEYEAFHCE GLCEFPLRSH LEPTNHAVIQ
TLMNSMDPES TPPTCCVPTR LSPISILFID SANNVVYKQY EDMVVESCGC R

Figure 1

```
                  10        20        30        40        50
MP 52  CSRKALHVNF KDMGWDDWII APLEYEAFHC HGLCEFPLRS HLEPTNHAVI
BMP 2  CKRHPLYVDF SDVGWNDWIV APPGYHAFYC HGECPFPLAD HLNSTNHAIV
BMP 4  CRRHSLYVDF SDVGWNDWIV APPGYQAFYC HGDCPFPLAD HLNSTNHAIV
BMP 5  CKKHELYVSF RDLGWQDWII APEGYAAFYC DGECSFPLNA HMNATNHAIV
BMP 6  CRKHELYVSF QDLGWQDWII APKGYAANYC DGECSFPLNA HMNATNHAIV
BMP 7  CKKHELYVSF RDLGWQDWII APEGYAAYYC EGECAFPLNS YMNATNHAIV
       * +  * * *  * ** ***+ **  * *+ * +* * *** + ++  ****

                  60        70        80        90        100
MP 52  QTLMNSMDPE STPPTCCVPT RLSPISILFI DSANNVVYKQ YEDMVVESCG  CR
BMP 2  QTLVNSVNS- KIPKACCVPT ELSAISMLYL DENEKVVLKN YQDMVVEGCG  CR
BMP 4  QTLVNSVNS- SIPKACCVPT ELSAISMLYL DEYDKVVLKN YQEMVVEGCG  CR
BMP 5  QTLVHLMFPD HVPKPCCAPT KLNAISVLYF DDSSNVILKK YRNMVVRSCG  CH
BMP 6  QTLVHLMNPE YVPKPCCAPT KLNAISVLYF DDNSNVILKK YRNMVVRACG  CH
BMP 7  QTLVHFINPE TVPKPCCAPT QLNAISVLYF DDSSNVILKK YRNMVVRACG  CH
       *** +++ ++ + *  **+**  *+ ** *   *   +*+ *  * +****++**  *+
```

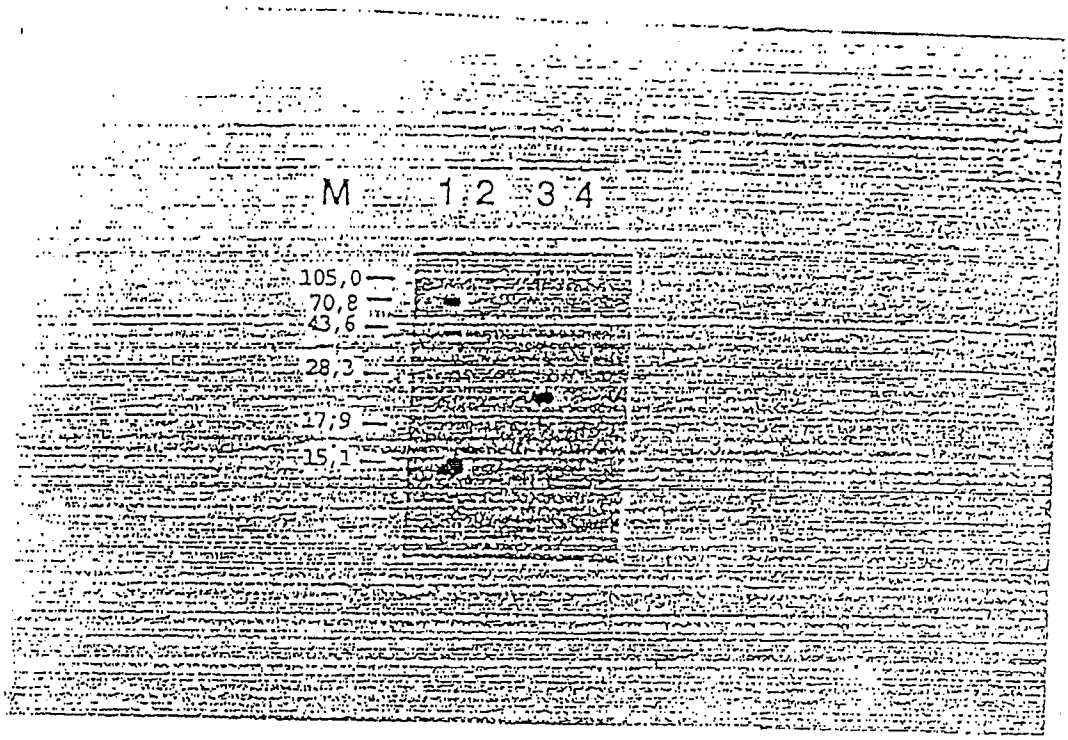
Figure 2a

	Eco RI Nco I
OD	ATGAATTCCCATGGACCTGGGCTGGMAKGAMTGGAT
BMP 2	ACGTGGGGTGGAATGACTGGAT
BMP 3	ATATTGGCTGGAGTGAATGGAT
BMP 4	ATGTGGGCTGGAATGACTGGAT
BMP 7	ACCTGGGCTGGCAGGACTGGAT
TGF-β1	AGGACCTCGGCTGGAAGTGGAT
TGF-β2	GGGATCTAGGGTGGAAATGGAT
TGF-β3	AGGATCTGGGCTGGAAGTGGGT
Inhibin α	AGCTGGGCTGGGAACGGTGGAT
Inhibin β_A	ACATCGGCTGGAATGACTGGAT
Inhibin β_B	TCATCGGCTGGAACGACTGGAT

Figure 2b

	Eco RI
OID	ATGAATTCGAGCTGGGTSGGSRCACAGCA
BMP 2	GAGTTCTGTGGGACACAGCA
BMP 3	CATCTTTTCTGGTACACAGCA
BMP 4	CAGTTCAGTGGGCACACAACA
BMP 7	GAGCTGGGTGGGGGCACAGCA
TGF-β1	CAGCGCCTGCGGCACGCAGCA
TGF-β2	TAAATCTTGGGACACGCAGCA
TGF-β3	CAGGTCCTGGGGCACGCAGCA
Inhibin α	CCCTGGGAGAGCAGCACAGCA
Inhibin β_A	CAGCTTGGTGGGCACACAGCA
Inhibin β_B	CAGCTTGGTGGGAATGCAGCA

Figure 3



M: prestained protein molecular weight marker with the stated apparent molecular weights listed (Gibco BRL #26041-020)

1: Cell culture supernatant (100 µl) after infection with recombinant viruses (with inserted MP52 cDNA) under reducing (1 % ß-mercaptoethanol) conditions

2: Cell culture supernatant (100 µl) after infection with wild-type viruses (without inserted foreign DNA) under reducing (1 % ß-mercaptoethanol) conditions

3: Cell culture supernatant (500 µl) after infection with recombinant viruses (with inserted MP52 cDNA) under non-reducing conditions

4: Cell culture supernatant (500 µl) after infection with wild-type viruses (without inserted foreign DNA) under non-reducing conditions